**Introduction**

The objective of this study of protein-protein interactions is to introduce a generic computational model to exploit the synergy among various protein information. The computational model is based on meta learning. It has the flexibility to integrate various component prediction tools into a unique prediction system.

For the purpose of demonstration, we provide the executable code of PPI-MetaGO. The user can input the Training Data and Test Data for PPI prediction.

**Availability**

The complete code and data is stored instead on the Google Drive at <https://drive.google.com/open?id=127bjTMPY68Q_VWpwhBPQoCYNOJ11JqIG> due to the limit of the file size on GitHub.

A simple demonstration code used for this study can be found at <https://github.com/mlbioinfolab/ppi-metago>.

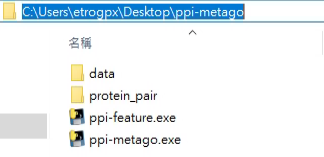
**Materials**

**Computer code:**

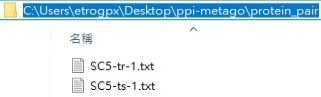
1. ppi-metago.exe: executable code of PPI-MetaGO
2. ppi-feature.exe: executable code of the feature generator that generates the feature values of protein pairs for training and testing.

**Data folders/subfolders:**

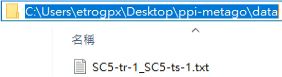
1. main\_folder: a folder to store the program ppi-metago.exe and data. The organization of a main folder may look like the snapshot below.



1. protein\_pair: a subfolder (named “protein\_pair”) that stores the files of protein pairs ppi-feature.exe requires to generate feature values. The snapshot below shows two files, one for training and the other for testing. The subfolder “protein\_pair” is the prespecified folder from which ppi-feature.exe reads a training data file, and a test data file. It generates the feature values, and produces a data file for training and testing ppi-metago.exe. (see next subsection).



1. data: a sub-folder (named “data”) that stores the data file produced by ppi-feature.exe for training PPI-MetaGO and predicting PPIs.



**Note.** All the folder/subfolder names are prespecified and cannot be changed.

**Instructions to use the meta classifiers to predict the protein protein interactions**

1. Create main folder “ppi-metago” and subfolders “protein\_pair” and “data” to store the code and the data. This main folder will serve as a working folder of PPI-MetaGO.
2. To complete a training and testing process, prepare one training data file and one test data file, both in text format. Place them in subfolder “protein\_pair”.

A sample training data file xxx.txt may look like the following. Each protein pair should be separated by a tab.

E.g. In xxx.txt file (a text file):

17799N 19113N

22289N 22550N

18381N 17383N

21607N 20929N

3. How to run PPI-MetaGO in Windows Systems:

1. Open the command line and switch to the ppi-metago folder.
2. Generate the feature values for training and testing by typing the following command,

ppi-feature.exe trfn=SC5-tr-1.txt tsfn=SC5-ts-1.txt

The first argument, trfn, specifies the training file, and the second, tsfn, specifies the test file.

If the protein sequence and GO term of a protein cannot be found in the database we provided, please add them into corresponding file manually.

The code “ppi-feature.exe” will read and process the files from the subfolder “protein\_pair” and outputs a feature file to the subfolder “data” folder.

The snapshot below shows an example.



1. To train PPI-MetaGO and predict protein-protein interactions, type the following command.

ppi-metago.exe fn=SC5-tr-1\_SC5-ts-1.txt

The following snapshot shows an example of the command.



1. When execution completed, ppi-metago.exe outputs a brief result summary on the screen, as shown in the snapshot below. A detailed result is also produced and stored in the subfolder “data.”

